

# A short review and primer on multimodal psychophysiological applications in work-related human computer interaction

Marco Filetti<sup>1</sup>, Jari Torniainen<sup>2</sup>

<sup>1</sup> Helsinki Institute for Information Technology HIIT, Department of Computer Science, Aalto University, Espoo, Finland

`marco.filetti@aalto.fi`

<sup>2</sup> Quantitative Employee unit, Finnish Institute of Occupational Health, Helsinki, Finland

**Abstract.** The application of psychophysiology in human-computer interaction is a growing field with significant potential for future smart personalised systems. Working in this emerging field requires comprehension of an array of physiological signals and analysis techniques. This paper focuses on the aggregation of multiple physiological measurements, obtained from one or more sensors. This approach requires the classification of relatively large samples of multidimensional data, which must be associated to specific cognitive or affective states. Researchers generally attempt to solve this problem by utilising machine learning techniques.

We present a short review to serve as a primer for the novice, enabling rapid familiarisation with the latest core concepts. We put special emphasis on work-related human-computer interface applications to distinguish from the more common clinical or sports uses of psychophysiology.

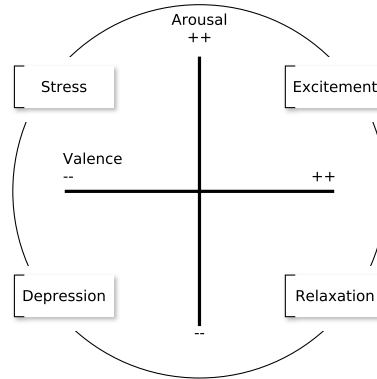
This paper is an extract from a comprehensive review of the entire field of ambulatory psychophysiology, including 12 similar chapters, plus application guidelines and systematic review. Thus any citation should be made using the following reference:

B. Cowley, M. Filetti, K. Lukander, J. Torniainen, A. Henelius, L. Ahonen, O. Barral, I. Kosunen, T. Valtonen, M. Huotilainen, N. Ravaja, G. Jacucci. *The Psychophysiology Primer: a guide to methods and a broad review with a focus on human-computer interaction*. Foundations and Trends in Human-Computer Interaction, vol. 9, no. 3-4, pp. 150–307, 2016.

**Keywords:** multimodal, psychophysiology, human-computer interaction, primer, review

## 1 Introduction

Signals obtained from multiple sources (physiological and behavioural) can be combined and analysed collectively for determining the state of a user (or, in



**Fig. 1.** The simple emotional circumplex model, with orthogonal bipolar dimensions of arousal (from alert to lethargic) and valence (pleasant to unpleasant).

some cases, multiple users). Multiple metrics can also be computed from a single signal. Using multiple measurements has the potential of providing more accurate estimates: while data collected from a single source could lead to conflicting interpretations, using data from multiple sources in combination aids in disambiguating cognitive or affective state. In the case of the widely known two-dimensional model of affect [Russell, 1980] (a simplified version of which is depicted in Figure 1), a moderate increase in arousal could be caused by the subject experiencing either fear or annoyance. For the purpose of our example, we assume that this increase in arousal is unequivocally measured via EDA (discussed in [Cowley et al., 2016, Section 3.2]). Integrating EDA with additional measurement (such as facial recognition) can assist in discriminating which of the two emotions the user was actually experiencing; for instance, the vertical position of the eyebrows is often higher when one is experiencing fear rather than annoyance.

Similarly, when facial recognition alone does not provide enough information to detect the emotion unambiguously, measuring arousal (from EDA or other signals) could aid in correctly classifying less clear-cut cases. For example, Bailenson et al. [2008] found males to be less facially expressive than females when experiencing sadness. Accordingly, including additional physiological signals in their classifier (such as EDA and ECG) increased the probability of correctly detecting sadness for males. Mandryk and Atkins [2007] presented an interesting approach, using a fuzzy classifier based on ECG, EDA, and facial EMG, to classify both arousal and valence of participants playing a popular computer game.

Novak et al. [2012] and Lisetti and Nasoz [2004] discuss general strategies that can be employed in the design and implementation of multimodal systems.

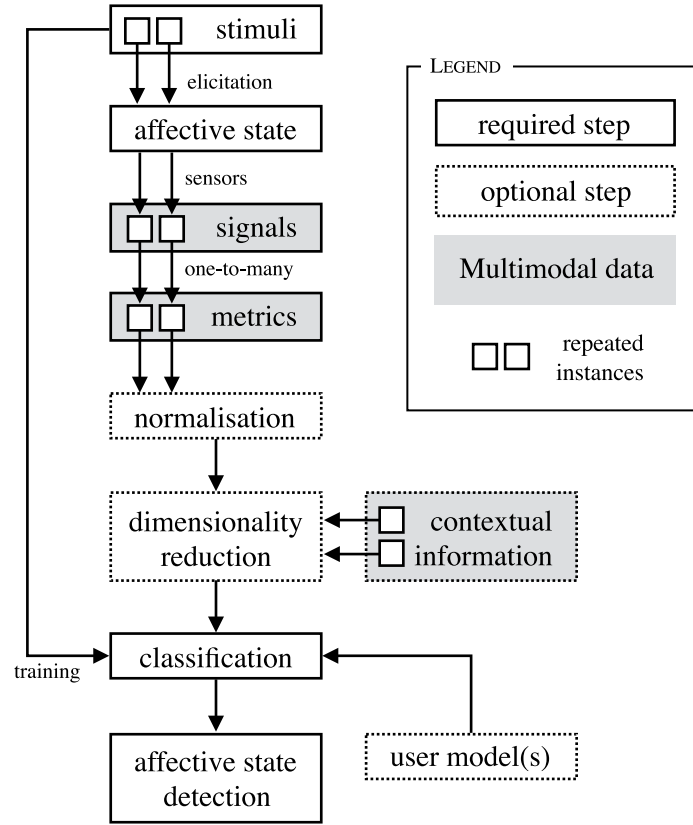
Inspired by these works, we describe here a generic schema for detecting affective and cognitive state in the types of multimodal systems discussed in this part of the primer; see Figure 2. As this figure shows, stimuli are associated with some cognitive/affective state(s) in the first step, either through the use of validated stimuli (e.g., the International Affective Picture System) or via subjective labelling of each stimulus after perception (for instance, a questionnaire can be used to evaluate a recording of the experience). Metrics are computed from their respective signals. Then, a classifier is trained to associate the data obtained with specific affective states on the basis of the stimulus classification performed earlier. ‘Optional’ procedures, shown in dotted boxes in the figure, may be applied at this stage. Such optional procedures may increase classification accuracy. For example, one may include a user model based on population-level inferences or (non-physiological) contextual information – such as time of day. Finally, reverse inference is performed to associate a particular set of data with a single specific state or with a range of states; probabilities are used to identify those states that are more likely to be associated with the observed data.

Multimodal systems can be useful when data from a given channel are missing, a situation that can normally be expected to arise when data are recorded outside research laboratories. For example, in the research by Wagner et al. [2011], missing data were handled by means of a naïve Bayes classifier and three modalities: voice (audio), facial recognition (video), and gestures (accelerometer).

**Determination of affective and cognitive state** Several studies have considered fusing information from multiple signals in order to assess mental workload. In a recent study, Hogervorst et al. [2014] noted that the use of multiple physiological signals is expected to enhance estimation of mental workload if the chosen signals represent separate aspects of workload. They extracted features from EEG, ECG, skin conductance, respiration, pupil size, and eye blinks, using these as inputs for both support vector machine (SVM) and elastic net classifiers. They achieved a high classification accuracy in two-level workload determination.

Cognitive state can be estimated also through a combination of a single physiological signal and a non-physiological source – Mühl et al. [2014], for example, used an LDA classifier to classify mental workload, combining EEG and mood (measured via self-reporting). Fusing of multiple features from the same signal has been undertaken, for instance, by Brouwer et al. [2012], who used features derived from EEG as inputs to an SVM classifier. They concluded that models using both EEG and ERP features – i.e., fusion models – worked better for mental workload classification using short data segments.

Given that learning can be facilitated by an optimal level of arousal [Baldi and Bucherelli, 2005, Sage and Bennett, 1973], Cowley and Ravaja [2014] showed that the interaction of EDA and EEG features predicted learning outcomes in a game-like task wherein individual signals alone were uninformative.



**Fig. 2.** Diagrammatic generalisation of multimodal physiological systems. Grey boxes are roughly equivalent to the concept of a ‘complex topography’ in signal processing systems (as described in [Cowley et al., 2016, Section 3.12]).

A study by Wilson and Russell [2003a] applied a combination of 85 EEG frequency power features, inter-beat interval, respiration rate, and electro-oculogram data from air traffic controllers. These 88 features were classified via an artificial neural network (ANN), achieving very high accuracy in two-class mental workload classification. Similar performance was achieved in classification of mental workload on a three-level scale instead [Wilson and Russell, 2003b]. The same set of signals were used by Wilson and Russell [2007] in an adaptive-aiding system.

**Applications** Multimodal systems capable of interpreting physiological signals in realistic HCI applications have been implemented with some success, especially in the last 10 years. We will discuss the most relevant applied work here.

Healey and Picard [2005] examined drivers’ stress levels (referring to distress, or negative stress) via four modalities: ECG, EMG, EDA, and respiration.

Using Fisher's discriminant analysis, they reported an accuracy of 97.4%. This relatively early research focused only on assessing the feasibility of correctly classifying three levels of stress. Malta et al. [2011] conducted a similar study, in which drivers' frustration was measured during actual (non-simulated) driving. Using a Bayesian network, they classified specific time segments in two classes (frustration-present vs. frustration-absent). They achieved, for the most part, 80% hit and 9% false positive rates, with the rates depending on the data fed to the classifier. Interestingly, the most accurate predictions were obtained when contextual information, such as traffic density, was provided to the classifier along with physiological information.

Pavlidis et al. [2007] utilised thermal cameras to measure multiple signals. Although only a single sensor was employed technically (the camera's thermal sensor), they effectively measured blood flow, cardiac pulse, and breathing rate. These metrics were used to compute a general level of negative stress, by means of a model of data fusion developed specifically for this application. The researchers reported good accuracy in detecting the actual stress levels ( $r = 0.91$ , Pearson correlation), with the exception of one outlier. They proposed two applications, desktop computer monitoring and sleep analysis, in both of which a stress level was computed for the current user, who was then alerted in the event that a specific threshold was exceeded. It should be noted, though, that the desktop computer test was carried out in the form of a laboratory experiment (a variant of the Stroop Test) [Pavlidis et al., 2007].

In other work, a biometric mouse and finger sensor (intended for desktop computer users) were used to measure EDA, skin temperature, heart rate, and touch intensity, in conjunction with behavioural data (such as mouse movements), for development of a recommender system that was tested in real-world working conditions over a span of four years [Kaklauskas et al., 2011]. The recommender system was capable of detecting stress and anxiety. It then presented individual users with suggestions for managing their work environment, on demand. Similar research involved a multimodal system employing EEG, ECG, and EMG, which was tested by CAD engineers. The system classified four emotions (frustration, challenge, engagement, and satisfaction) in users carrying out CAD work (so that, for example, one could detect designs developed amid frustration, which might display errors in judgement). However, small sample sizes prevented statistically significant findings [Liu et al., 2013]. In other applications, it was noted that less expert users might benefit from physiological integration, as in a system that utilises gaze, facial recognition, and speech recognition in order to predict user intentions. However, while welcomed by non-experts, a system of this nature [Maat and Pantic, 2007] was met with dissatisfaction by more experienced users.

Measurement of physiology from multiple sources has not been limited to single-user desktop computer usage. It has been carried out in more social settings also, such as at meetings or on public speaking occasions. For example, stress-related arousal has been measured 'in the wild' through a combination of EDA, ECG, and motion sensors (though the system was tested with only a

small number of participants) [Kusserow et al., 2013]. This creates the possibility of analysing one's own data during a public speaking event, so that any performance problems due to distress can be identified and, possibly, be avoided in the future. In addition, observer stress, the stress experienced by a person observing a meeting or group event, has been measured by means of thermal cameras, EDA, and EEG [Sharma and Gedeon, 2014]. In the same study, observer stress was measured during an interview event (with the interview being designed to induce observer stress at specific moments). With genetic algorithms and support vector machines, the presence (or absence) of observer stress was detectable with an accuracy of 98%. In a similar context, a more laboratory-oriented setting (which used virtual humans as 'presenters') utilised video cameras (for facial recognition), eye tracking, and audio signals to measure observer interest in a given topic [Schuller et al., 2009]. That study demonstrated that a virtual agent was capable of detecting loss of the user's interest and of switching topic accordingly. Although the frequent change of topic might have had an adverse impact on user understanding, the system was capable of correctly detecting changes in user interest (with a cross-correlation value of 0.72, using support vector regression).

**Challenges** As are all the other methods mentioned, multimodal methods are affected by high noise levels at the receiving end. The 'bleeding-edge' nature of the technologies employed implies that signal quality can vary greatly across devices. Moreover, the interference typical of body-based measurements (such as adverse effects of muscle movements on EEG) complicates processing of the signals obtained. In general, this means that 'one size fits all' measurements are infeasible, so user- and environment-based customisations are required for obtaining satisfactory signal quality, especially when physiology is being measured outside well-controlled laboratory settings [Hong and Dey, 2014].

If they are to be effective, multimodal measurements would benefit from an overarching theory of emotion. Although some approaches pointing in this direction have been put forward [Conati and Maclaren, 2009, Peter and Herbon, 2006, Gunes and Schuller, 2013, Gunes and Pantic, 2010], consensus is still lacking, as the nature of emotions is not fully understood. Moreover, a lack of interdisciplinary work is evident in this field: applications devised within one community are rarely shared with other communities (for example, an HCI experiment may not demonstrate strong enough correlations between the data and specific events to satisfy the demands of psychological research, while psychological research is usually too theoretically oriented for HCI applications) [Lopatovska and Arapakis, 2011].

The absence of consensus with regard to theories of emotion might also explain why multimodal methods generally manifest only modest improvements over single-sensor settings [D'Mello and Kory, 2012]. Often, multimodal systems assume co-ordinated bodily responses to a single emotion, which might not occur in reality. Since multimodal measurements demand more computing power, it is important to verify that multiple theories of emotion are considered while the

work is still in the development stage, so as to avoid the creation of a highly complex system that yields minimal gains over a single-sensor-based system. That said, modest gains are still welcome, provided that the increase in computation time is not unmanageable. Moreover, development of novel systems that take current theories of emotion into account could lead to new insights into the nature and physiology of emotions themselves (e.g., in the case of unexpected findings).

**Conclusion** Most of the multimodal applications described here, especially in the context of HCI, have focused on detection of (negative) stress, possibly because stress is often identified as the most negative – yet measurable – factor that can arise in working environments. Applications in this area have found some success, although they rarely address use cases beyond user self-assessment. Multimodal applications that constitute attempts to detect other cognitive or affective states, such as interest, intention, or emotion, do exist but are only in the early stages of their development. For that reason, research in this connection is currently focused on improving classification accuracies rather than on presenting effective use cases.

It appears that research into multimodal physiological systems would benefit from more interdisciplinary work focused on finding common ground between the fields of psychology and human–computer interaction, so that advanced systems, capable of detecting user states with greater accuracy, can be developed.

## Bibliography

- Jeremy N. Bailenson, Emmanuel D. Pontikakis, Iris B. Mauss, James J. Gross, Maria E. Jabon, Cendri A.C. Hutcherson, Clifford Nass, and Oliver John. Real-time classification of evoked emotions using facial feature tracking and physiological responses. *International Journal of Human-Computer Studies*, 66(5):303–317, 2008.
- Elisabetta Baldi and Corrado Bucherelli. The inverted ‘u-shaped’ dose-effect relationships in learning and memory: Modulation of arousal and consolidation. *Nonlinearity in biology, toxicology, medicine*, 3(1):9–21, 2005. ISSN 1540-1421.
- Anne-Marie Brouwer, Maarten A Hogervorst, Jan BF van Erp, Tobias Heffelaar, Patrick H Zimmerman, and Robert Oostenveld. Estimating workload using EEG spectral power and ERPs in the n-back task. *Journal of neural engineering*, 9(4):045008, 2012.
- Cristina Conati and Heather Maclaren. Empirically building and evaluating a probabilistic model of user affect. *User Model User-Adap Inter*, 19(3):267–303, 2009.
- Ben Cowley and Niklas Ravaja. Learning in Balance: Using Oscillatory EEG Biomarkers of Attention, Motivation and Vigilance to Interpret Game-Based Learning. *Cogent Education*, 1(1):1–23, 2014. ISSN 2331-186X.
- Benjamin Cowley, Marco Filetti, Kristian Lukander, Jari Torniaainen, Andreas Henelius, Lauri Ahonen, Oswald Barral, Ilkka Kosunen, Teppo Valtanen, Minna Huottilainen, Niklas Ravaja, and Giulio Jaccuci. The Psychophysiology Primer: a guide to methods and a broad review with a focus on human computer interaction. *Foundations and Trends in HCI*, 9(3-4):150–307, 2016.
- Sidney D’Mello and Jacqueline Kory. Consistent but modest: A meta-analysis on unimodal and multimodal affect detection accuracies from 30 studies. In *Consistent but modest: A meta-analysis on unimodal and multimodal affect detection accuracies from 30 studies*, pages 31–38. ACM, 2012.
- Hatice Gunes and Maja Pantic. Automatic, dimensional and continuous emotion recognition. *International Journal of Synthetic Emotions (IJSE)*, 1(1):68–99, 2010.
- Hatice Gunes and Björn Schuller. Categorical and dimensional affect analysis in continuous input: Current trends and future directions. *Image and Vision Computing*, 31(2):120–136, 2013.
- J.A. Healey and R.W. Picard. Detecting stress during real-world driving tasks using physiological sensors. *IEEE Trans. Intell. Transport. Syst.*, 6(2):156–166, 2005.
- Maarten A Hogervorst, Anne-Marie Brouwer, and Jan BF van Erp. Combining and comparing eeg, peripheral physiology and eye-related measures for the assessment of mental workload. *Frontiers in neuroscience*, 8, 2014.
- Jin-Hyuk Hong and Anind K Dey. Affect modeling with field-based physiological responses. *Interacting with Computers*, page iwu014, 2014.



- Arturas Kaklauskas, Edmundas Kazimieras Zavadskas, Marko Seniut, Gintautas Dzemyda, V Stankevic, C Simkevicius, T Stankevic, R Paliskiene, A Matuliaskaite, and S Kildiene. Web-based biometric computer mouse advisory system to analyze a user's emotions and work productivity. *Engineering Applications of Artificial Intelligence*, 24(6):928–945, 2011.
- Martin Kusserow, Oliver Amft, and Gerhard Tröster. Monitoring stress arousal in the wild. *IEEE Pervasive Computing*, 12(2):28–37, 2013.
- Christine L Lisetti and Fatma Nasoz. Using noninvasive wearable computers to recognize human emotions from physiological signals. *EURASIP Journal on Applied Signal Processing*, 2004:1672–1687, 2004.
- Y Liu, JM Ritchie, T Lim, Z Kosmadoudi, A Sivanathan, and RCW Sung. A fuzzy psycho-physiological approach to enable the understanding of an engineer's affect status during cad activities. *Computer-Aided Design*, 2013.
- Irene Lopatovska and Ioannis Arapakis. Theories, methods and current research on emotions in library and information science, information retrieval and human–computer interaction. *Information Processing & Management*, 47(4):575–592, 2011.
- Ludo Maat and Maja Pantic. Gaze-x: Adaptive, affective, multimodal interface for single-user office scenarios. In Thomas S. Huang, Anton Nijholt, Maja Pantic, and Alex Pentland, editors, *Artificial Intelligence for Human Computing: ICMI 2006 and IJCAI 2007 International Workshops, Banff, Canada, November 3, 2006, Hyderabad, India, January 6, 2007, Revised Selected and Invited Papers*, pages 251–271. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007. ISBN 978-3-540-72348-6.
- Lucas Malta, Chiyomi Miyajima, Norihide Kitaoka, and Kazuya Takeda. Analysis of real-world driver's frustration. *Intelligent Transportation Systems, IEEE Transactions on*, 12(1):109–118, 2011.
- Regan L Mandryk and M Stella Atkins. A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. *International journal of human-computer studies.*, 65(4):329–347, 2007.
- Christian Mühl, Camille Jeunet, and Fabien Lotte. Eeg-based workload estimation across affective contexts. *Frontiers in neuroscience*, 8, 2014.
- Domen Novak, Matjaž Mihelj, and Marko Munih. A survey of methods for data fusion and system adaptation using autonomic nervous system responses in physiological computing. *Interacting with Computers*, 24(3):154–172, 2012.
- Ioannis Pavlidis, Jonathan Dowdall, Nanfei Sun, Colin Puri, Jin Fei, and Marc Garbey. Interacting with human physiology. *Computer Vision and Image Understanding*, 108(1):150–170, 2007.
- Christian Peter and Antje Herbon. Emotion representation and physiology assignments in digital systems. *Interacting with Computers*, 18(2):139–170, 2006.
- James A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178, 1980.
- George H. Sage and Bonnie Bennett. The Effects of Induced Arousal on Learning and Performance of a Pursuit Motor Skill. *Research Quarterly. American Association for Health, Physical Education and Recreation*, 44(2):140–149, 1973. ISSN 1067-1188.

- Björn Schuller, Ronald Müller, Florian Eyben, Jürgen Gast, Benedikt Hörnler, Martin Wöllmer, Gerhard Rigoll, Anja Höthker, and Hitoshi Konosu. Being bored? recognising natural interest by extensive audiovisual integration for real-life application. *Image and Vision Computing*, 27(12):1760–1774, 2009.
- Nandita Sharma and Tom Gedeon. Modeling observer stress for typical real environments. *Expert Systems with Applications*, 41(5):2231–2238, 2014.
- Johannes Wagner, Elisabeth Andre, Florian Lingenfelser, and Jonghwa Kim. Exploring fusion methods for multimodal emotion recognition with missing data. *Affective Computing, IEEE Transactions on*, 2(4):206–218, 2011.
- Glenn F Wilson and Chris A Russell. Operator functional state classification using multiple psychophysiological features in an air traffic control task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 45(3):381–389, 2003a.
- Glenn F Wilson and Christopher A Russell. Real-time assessment of mental workload using psychophysiological measures and artificial neural networks. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 45(4):635–644, 2003b.
- Glenn F Wilson and Christopher A Russell. Performance enhancement in an uninhabited air vehicle task using psychophysiological determined adaptive aiding. *Human factors: The journal of the human factors and ergonomics society*, 49(6):1005–1018, 2007.